



12-22-2003

Marginalized Transition Models for Longitudinal Binary Data With Ignorable and Nonignorable Dropout

Brenda F. Kurland

University of Washington, bkurland@fhcrc.org

Patrick J. Heagerty

University of Washington, heagerty@u.washington.edu

Suggested Citation

Kurland, Brenda F. and Heagerty, Patrick J., "Marginalized Transition Models for Longitudinal Binary Data With Ignorable and Nonignorable Dropout" (December 2003). *UW Biostatistics Working Paper Series*. Working Paper 222.
<http://biostats.bepress.com/uwbiostat/paper222>

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

Copyright © 2011 by the authors

1. INTRODUCTION

In a longitudinal evaluation comparing three pharmacological treatments for schizophrenia,¹ 431 subjects were randomly assigned to treatment groups and 420 completed at least one post-randomization evaluation. The Positive and Negative Syndrome Scale (PANSS),² an established instrument for measuring severity of schizophrenia, was administered at selection into the trial, after the washout period, and after 1,2,4,6, and 8 weeks of treatment. Diggle, Heagerty, Liang and Zeger³ analyzed these 7 total PANSS scores to illustrate regression models for continuous response data subject to dropout. Published clinical results^{1,4} used a binary primary outcome: clinically significant improvement, measured as a 20% reduction in PANSS compared to baseline. In this manuscript, we consider clinically significant improvement as the study response, and examine three treatment groups over the sequence of 5 post-baseline scores.

Relevance of the dropout process in modeling PANSS responses is apparent in Figure 1, which shows the proportion of participants showing clinical improvement separately for those who drop out after 1, 2, 4, or 6 weeks of treatment. Patients who remain in the study for the entire 8 weeks are more likely to show improvement at almost every timepoint. Furthermore, the proportion of subjects showing improvement appears to decrease in the assessment preceding dropout.

Classes of models to accommodate longitudinal data with dropout are summarized by Diggle et al.³ *Selection models*⁵ factorize the joint density of dropout and response information into dropouts given responses, and responses. *Pattern-mixture models*⁶ factorize the joint density with response models conditioning on dropout information. In *shared random effects models*,⁷ response and dropout variables may be conditionally independent given random effects values. Figure 1 illustrates a pattern-mixture approach, since response profiles are stratified by dropout time. Pattern-mixture

results, stratified by length of followup, would not directly be useful for making clinical decisions about treatment efficacy. Random effects methods' regression models will be subject-specific⁸ when the response is binary. For treatment comparisons without conditioning on random effects or dropout patterns, we therefore focus on selection models to accommodate longitudinal binary data with monotone dropout.

Likelihood-based methods proposed for analysis of binary longitudinal data with nonignorable dropout have had limitations of computation, interpretability, and modeling assumptions. Fitzmaurice, Laird and Zahner⁹ and Molenberghs, Kenward and Lesaffre¹⁰ proposed selection models combining a marginal regression model with an association model based on, respectively, conditional and marginal odds ratios. The likelihood formulation of Fitzmaurice et al.⁹ requires balanced complete data and is not reproducible: if cluster sizes change, the specification and interpretation of the association model changes. For both methods, the number of parameters needed to model association among responses grows rapidly as the cluster size increases. Although sensitivity analysis is recommended, missingness model parameters are estimated even when weak identifiability is dependent on distributional assumptions of the model.¹¹

Regression parameters can be estimated consistently by generalized estimating equations¹² for MAR or NI missing data (defined in Section 2.1) by weighting the estimating equations by inverse probability of sampling weights.^{13–15} These weights allow observed responses to stand as proxies for missing responses of subjects with similar characteristics. Parametric models to compute weights are similar to missingness models for likelihood-based selection models. If weights are modeled incorrectly, regression parameters will not be consistent. Semiparametric models will be less efficient than their likelihood-based counterparts, unless the association modeled is very close to the true

association model.¹⁶

Multiple imputation (MI)¹⁷ may be used to analyze longitudinal data with missing values. Bayesian analysis of longitudinal binary data may also be extended to accommodate missing values.¹⁸ Due to asymptotic equivalence between Bayesian and likelihood-based methods, appropriately specified MI and maximum likelihood methods will yield similar parameter estimates and standard errors (for models with a large sample size and sufficient number of imputations).^{17,19} Paik²⁰ proposed semiparametric sequential imputation to find consistent estimates of mean parameters under MAR and NI dropout. For saturated mean and selection models, Paik's mean imputation method yields identical results to the weighted GEE of Robins, Rotnitzky and Zhao.¹³

These existing methods for analysis of binary longitudinal data with ignorable or nonignorable dropout differ with respect to target of inference, robustness to dropout misspecification, and ease of computation. The marginalized transition model,²¹ extended here to accommodate NI dropout, has useful properties in these three areas.

Heagerty and Zeger²² described a class of *marginalized models* in which a conditional model (such as a random effects or transition model) specifies the underlying process of data generation, but marginal mean parameters may be estimated. The marginalized transition model,²¹ described below, retains an association model based on conditional expectations (transitions), but permits direct marginal regression estimation of treatment effects. For other transition models for longitudinal binary data,²³ treatment effects are conditional on values of prior responses. The marginalized transition model (MTM), in which association is modeled via conditional dependence on prior responses, and the selection model, in which dropout is related to responses, provide a coherent likelihood-based framework for describing both response and dropout processes.

The major goals of this manuscript are: to describe how the MTM provides a computationally feasible framework for analysis under MCAR, MAR, and NI dropout; and to compare the efficiency and potential for bias of a MTM likelihood approach and an inverse probability of censoring weighted GEE (IPCW-GEE) approach. In Section 2 we extend the marginalized transition model to accommodate monotone NI missing data models. Section 3 compares the MTM and IPCW-GEE with respect to efficiency and misspecification bias, and Section 4 compares the MTM and IPCW-GEE in analysis of the PANSS data introduced above.

2. MARGINALIZED TRANSITION MODEL WITH DROPOUT

A likelihood-based method for analysis of serial binary data with marginal mean parameters and first-order Markov association was introduced by Azzalini²⁴ and generalized as a marginalized transition model by Heagerty.²¹ This section describes the models and the likelihood function for first- and second-order marginalized transition models with both ignorable and nonignorable monotone dropout.

2.1. Notation

Some notation is necessary to describe methods for analyzing binary longitudinal data with dropout. Each response vector, $\mathbf{Y}_i = (Y_{i1} \dots Y_{in_i})$ for individuals $i = 1, \dots, N$ is a vector of 0/1 values. For the PANSS data, each response Y_{it} reflects whether or not subject i showed improvement at time t , compared to baseline. The relationship between \mathbf{Y}_i and covariates is described by the $(n_i \times p)$ design matrix \mathbf{X}_i and the regression parameter p -vector β . We restrict ourselves to logistic regression:

$$\text{logit}(\mu_{it}^M) = \mathbf{X}_{it}\beta \tag{1}$$

where marginal mean $\mu_{it}^M = E(Y_{it}|X_{it})$ is the probability of positive response for subject i at time t , given covariate vector X_{it} . Serial correlation among members of vector Y_i is parameterized by vector α . The functional form for association (correlation) models and estimation of α differ for the different methods considered.

We consider only monotone dropout: if Y_{it} is missing, then Y_{is} is missing for $s \geq t$. The dropout process is described by a retention vector $R_i = (R_{i1} \dots R_{in_i})$, where R_{it} is an indicator taking value 1 if observation Y_{it} is observed and 0 if Y_{it} is missing. R_i is modeled by the $(n_i \times r)$ design matrix for dropout, D_i , and parameter r-vector ϕ . D_i may include responses (Y_i), covariates in X_i , and covariates not in X_i . An additional design matrix and parameter vector (G_i and λ) are introduced below for certain dropout models. Dropout parameter vector ϕ is assumed to be distinct from regression and association parameter vectors β and α . Response vector Y_i can be divided into two vectors, based on whether values are observed or missing: $Y_i = (Y_i^o, Y_i^m)$.

The hierarchy attributed to Little and Rubin²⁵ assesses the influence of missing values on estimation of parameters characterizing the response model. If dropout indicator R_i does not depend on any element of the response Y_i , the missingness mechanism is described as *missing completely at random* (MCAR). If dropout depends on observed (Y_i^o), but not unobserved (Y_i^m), responses, data are *missing at random* (MAR). If the dropout mechanism is dependent on unobserved response values Y_i^m (with or without dependence on Y_i^o), missingness is *nonignorable* (NI). If β is to be found by maximum likelihood and the missingness mechanism is MCAR or MAR, the missingness is *ignorable* and a likelihood-based analysis of available data will yield consistent estimates for β .²⁶ For generalized estimating equations (GEE),^{12,13} β will be consistent in an available case analysis if data are MCAR, but not if data are MAR or NI.¹²

2.2. Marginalized Transition Model

The marginalized transition model (MTM)²¹ is constructed around two models. Regression structure is characterized using a generalized linear model for the marginal mean, and serial dependence is captured by specifying conditional dependence of current outcomes on past outcomes. The association (serial dependence) model is essentially a Markov, or *transition* model,³ with parameters α that determine the strength of association between past outcomes and the current response. The innovation in the MTM approach is the ability to use conditional mean models for dependence while preserving the ability to structure the marginal mean directly, using a regression model for $E(Y_{it}|X_{it})$ indexed by a parameter vector β .

Figure 2, adapted from Figure 1 in Heagerty,²¹ is a schematic representation of serial observations that follow a second-order marginalized transition model (MTM) subject to nonignorable (NI) monotone dropout. Figure 2 illustrates the marginal mean, dependence of Y_{it} on past responses, and dependence of R_{it} on past values of R_i , past values of Y_i , and (if dropout is NI) on Y_{it} . Dashed boxes and parameters β represent the marginal mean that is assumed to be the primary target of inference. The conditional mean, in which the response at time t , Y_{it} , depends on covariates and prior responses, is parameterized by Δ and α . Finally, the monotone dropout process $f(R_i|Y_i)$ is parameterized by ϕ . Likelihood-based inference requires full specification of $f(Y_i, R_i)$, which is provided by the MTM and selection model described below.

2.3. Marginalized Transition Model Likelihood

When the dropout process $f(R_i|Y_i)$ is ignorable (MCAR or MAR), the likelihood function for marginal mean parameters β (Equation 1) and conditional mean (serial dependence) parameters α

can be defined as follows for response vectors Y_i , for clusters $i = 1, \dots, N$ at timepoints $t = 1, \dots, n_i$:

$$\begin{aligned} \mathbf{L}_i(\beta, \alpha) &= P(Y_{i1} = y_{i1}) \cdot P(Y_{i2} = y_{i2} | Y_{i1}) \cdots P(Y_{in_i} = y_{in_i} | Y_{i1} \dots Y_{in_i-1}) \\ &= \prod_{i=1}^N (\mu_{i1}^M)^{y_{i1}} \cdot (1 - \mu_{i1}^M)^{(1-y_{i1})} \prod_{i=1}^N \prod_{t=2}^{n_i} (\mu_{it}^C)^{y_{it}} \cdot (1 - \mu_{it}^C)^{(1-y_{it})}. \end{aligned}$$

For a p^{th} -order marginalized transition model, $\text{MTM}(p)$, the logit-linear model for μ_{it}^C is:

$$\begin{aligned} \text{logit}(\mu_{it}^C) &= \Delta_{it} + \sum_{j=1}^p \gamma_{it,j} \cdot y_{it-j} \\ \text{for } \gamma_{it,j} &= Z_{it,j} \alpha_j. \end{aligned} \tag{2}$$

The association is modeled through p α -vectors. For example, if each $Z_{it,j} = 1$, then $\gamma_{it,j} = \alpha_j$, and $(\alpha_1 \dots \alpha_p)$ are each constants measuring the strength of association between the j th lagged response and the present response. Interactions of lagged response variables are possible in the model for the conditional mean, but they are not included in this presentation. Intercept Δ_{it} describes the conditional mean when the p previous response values are zero. For a given β (Equation 1) and α (Equation 2), the value of Δ_{it} may be derived through relationships between marginal and conditional means, as shown in the following convolution equation:

$$\begin{aligned} \mu_{it}^M &= \sum_{y_{it-1}, \dots, y_{it-p}} P(Y_{it} = 1 | Y_{it-1} \dots Y_{it-p}) \cdot P(Y_{it-1} = y_{it-1}, \dots, Y_{it-p} = y_{it-p}) \\ &= \sum_{y_{it-1}, \dots, y_{it-p}} \mu_{it}^C \cdot P(Y_{it-1} = y_{it-1}, \dots, Y_{it-p} = y_{it-p}). \end{aligned}$$

Further details are given by Heagerty.²¹

2.4. MTM Likelihood with Nonignorable Dropout

The likelihood function for regression, association, and dropout parameters (β, α, ϕ) is based on observed data: response data for subjects who have not yet dropped out (Y_i^o), and dropout indicator

(R_i) and covariate matrix \mathbf{X}_i for all time periods. Data are assumed to be generated under the complete data distribution $f(Y_i, R_i)$, where $Y_i = (Y_i^o, Y_i^m)$, and Y_i^m is a vector of responses that are missing. The observed data likelihood is:

$$\begin{aligned} L_i(\beta, \alpha, \phi) \propto f(Y_i^o, R_i) &= \int_{Y^m} f(Y_i, R_i) = \int_{Y^m} f(R_i | Y_i) f(Y_i) \\ &= \int f(R_i | Y_i) f(Y_i^o) f(Y_i^m | Y_i^o) d(Y_i^m | Y_i^o). \end{aligned} \quad (3)$$

This is a typical selection model factorization.^{25,5}

We now model $f(R_i | Y_i)$ in Equation 3 by logistic regression. First, use a telescoping product of conditional distributions to specify vector R_i :

$$\begin{aligned} f(R_i | Y_i) &= f(R_{i1} | Y_i) f(R_{i2} | R_{i1}, Y_i) \cdots f(R_{in_i} | H_{in_i}^{(R)}, Y_i) \\ &= \prod_{t=2}^{n_i} f(R_{it} | H_{it}^{(R)}, H_{it}^{(Y)}, Y_{it}) \end{aligned}$$

where $H_{it}^{(R)}$ is the history for R_i through time $(t-1)$, $H_{it}^{(R)} = (R_{i1} \dots R_{it-1})$, and $H_{it}^{(Y)}$ is the history for response Y_{it} in cluster (subject) i . The first observation, Y_{i1} , is assumed never to be missing: this common assumption^{5,13} simplifies parameterization of dropout models. Current values R_{it} and Y_{it} are assumed to depend only on past and current values, not the future.

Let $u_{it} = P(R_{it} = 1 | H_{it}^{(R)}, H_{it}^{(Y)}, Y_{it})$ and define time d as the first time that response Y is missing. ($R_{id-1} = 1$ and $R_{id} = 0$.) The likelihood contribution for each $R_{it} = r_{it}$ given the response history is:

$$f(R_{it} | H_{it}^{(R)}, H_{it}^{(Y)}, Y_{it}) = u_{it}^{r_{it}} (1 - u_{it})^{1-r_{it}}$$

where

$$u_{it} = \begin{cases} p_{it} & : t \leq d \\ 0 & : t > d \end{cases}$$

and

$$p_{it} = P(R_{it} = 1 | R_{t-1} = 1, H_{it}^{(Y)}, Y_{it}).$$

Parameterization of p_{it} is described below. The likelihood contribution of $f(R_i | Y_i)$ is:

$$\begin{aligned} f(R_i | Y_i) &= \prod_{t=2}^{d-1} p_{it}^{r_{it}} (1 - p_{id})^{1-r_{id}} \prod_{k=d+1}^n (1)^{(1-r_{ik})} \\ &= \left(\prod_{t=2}^{d-1} p_{it} \right) (1 - p_{id}). \end{aligned} \quad (4)$$

Substituting, and rearranging terms from Equations 3 and 4, the full likelihood may be written as:

$$\mathbf{L}_i(\beta, \alpha, \phi) = \underbrace{f(Y_i^o)}_{L1} \underbrace{\left(\prod_{t=2}^{d-1} p_{it} \right)}_{L2} \underbrace{\int (1 - p_{id}) f(Y_i^m | Y_i^o) d(Y_i^m | Y_i^o)}_{L3}. \quad (5)$$

L1, L2, and L3 correspond to Equations 23-25 of Diggle and Kenward.⁵ L1 is the likelihood for the observed responses, the marginalized transition model (MTM) under ignorable dropout. L2 is defined by the dropout model, and L3 is the expected value of $(1 - p_{id})$ under the distribution $f(Y_i^m | Y_i^o)$.

L2 models p_{it} , the probability that Y_{it} is observed given the response history and covariate values. Monotone dropout in discrete time can be analyzed as discrete-time survival analysis using logistic regression.²⁷ A general model for dropout is:

$$\text{logit}(p_{it}) = D_{it}\phi + G_{it}\lambda \cdot Y_{it}.$$

Here \mathbf{D}_i is a model matrix that may include covariates and lagged observations, with ϕ as the corresponding parameters. \mathbf{G}_i is the model matrix and λ the parameter vector associated with the current observation, Y_{it} . Dropout is MAR if λ is a zero-vector, and MCAR if λ is a zero-vector and \mathbf{D}_i does not contain any response (Y_i) values. Vector λ is separated from ϕ because of identifiability issues discussed below.

L3 in Equation 5 is defined in terms of p_{id} and $f(Y_i^m|Y_i^o)$. Probability p_{id} is determined by the selection model described for L2, and parameterized by ϕ and λ . Because p_{id} only depends on Y_i^m through Y_{id} , L3 involves the distribution of Y_{id} given the previous response values, Y_i^o . Conveniently, this transition is already modeled by the marginalized transition model:

$$\begin{aligned} E_{Y^m|Y^o}(1 - p_{id}) &= \sum_{Y_{id}=0,1} (1 - p_{id}) \cdot P(Y_{id} = y_{id} | H_{id}^{(Y)}) \\ &= (1 - p_{id|Y_{id}=1}) \cdot P(Y_{id} = 1 | H_{id}^{(Y)}) + (1 - p_{id|Y_{id}=0}) \cdot P(Y_{id} = 0 | H_{id}^{(Y)}) \\ &= \frac{1}{1 + \exp(D_{id}\phi + G_{id}\lambda)} \cdot \mu_{id}^C + \frac{1}{1 + \exp(D_{id}\phi)} \cdot (1 - \mu_{id}^C). \end{aligned}$$

Note that L1 only contributes to estimation of (β, α) , L2 only contributes to estimation of ϕ , and L3 contributes to all three parameter vectors. However, if the dropout process is MAR ($\lambda \equiv 0$), then L3 reduces to

$$\frac{1}{1 + \exp(D_{id}\phi)},$$

which depends only on ϕ .

The first- and second-order MTM likelihoods are maximized by Newton-Raphson estimation for all parameters. Second derivatives for the MTM(1) likelihood function are computed analytically. For MTM(2), numerical second derivatives are computed as local slopes from small perturbations of parameter estimates.²⁸ Computations for maximizing the MTM likelihood are performed using customized code written in C and S-Plus. A detailed description of the algorithm is given by Kurland.²⁹ Parameter standard errors are estimated as the inverse of the observed information matrix. Nested marginalized transition models can be compared by likelihood ratio tests.

Parameters λ are effectively unidentifiable because response Y_{it} is not observed when $R_{it} = 0$. Although parameters may appear estimable, weak identifiability is due to distributional assump-

tions of the model.¹¹ Identifiability may be addressed by pooling information from a separate source of cross-sectional data into the likelihood³⁰ or through restrictions to the structure of the dropout model.²⁶ A conservative and practical approach, though, is to perform the analysis repeatedly with parameters λ held constant through a series of plausible values. This sensitivity analysis examines the impact of the fixed values on the remainder of the model.^{14,11}

The MTM's use of a selection model to accommodate monotone dropout leads to a natural separation between dropout and response processes, especially for data missing at random (MAR). Also, as described by Heagerty,²¹ computational time for the MTM increases linearly with cluster size, while for other likelihood-based methods,^{9,10} computational time increases exponentially with cluster size.

3. SIMULATION STUDIES

For the likelihood-based MTM method, properly specified regression parameter estimates will be more efficient than IPCW-GEE estimates if the association induced by the transition model is closer to the true association model than the IPCW-GEE working correlation. For binary responses, Fitzmaurice, Laird and Rotnitzky¹⁶ found that asymptotic relative efficiency of GEE regression parameters is often quite high, but may decrease sharply with increasing marginal correlation. Bias in MTM estimates may result even for a correctly specified regression model when data are MAR and the association model is incorrect, or when data are NI and association or missingness is incorrectly specified. IPCW-GEE may yield biased estimates when the missingness model is incorrect.³¹ This section uses simulation studies to explore relative efficiency and the magnitude of misspecification bias for the MTM and IPCW-GEE.

In general it is difficult to compare the modeling burden of a likelihood-based method (MTM) to a semiparametric approach (IPCW-GEE). However, when data are MAR the MTM and IPCW-GEE are operationally quite parallel. The MTM is based on two models: the primary marginal regression model and the secondary dependence model. Dependence is modeled by logistic regression of the current outcome conditional on the response history and covariates. IPCW-GEE also requires two models: the primary regression model, and a selection model for dropout. Like MTM dependence, the IPCW-GEE selection model is a logistic regression on the past outcomes and covariates; however, the response is the dropout indicator rather than the current outcome. Under MAR, the requisite *correct modeling of response* (for MTM) is quite similar to the requisite *correct modeling of dropout* (for IPCW-GEE). Finally, when data are NI the likelihood approach must also explicitly specify a selection model.

Longitudinal binary data with dropout are simulated via a second-order marginalized transition model. The regression model is chosen to represent two treatments compared over time. Response Y_{it} may be interpreted as the presence ($Y_{it} = 1$) or absence ($Y_{it} = 0$) of an undesirable symptom for subject i at time t . The setup for the simulations is summarized in Table 1. The `time` vector ends in 0 to facilitate the common scenario where scientific interest is in end-of-study group comparisons. The expected probability of the symptom is approximately 10% at `time` = -4 for both control (`group` = 0) and treatment (`group` = 1) groups, and rises to about 30% for the treatment group and to 60% for the control group at `time` = 0. For clusters of size 5, the MTM(2) serial dependence produces correlations among observations that are similar to an autoregression structure, $\text{corr}(Y_{ij}, Y_{ik}) \approx \rho^{|j-k|}$ with $\rho \approx 0.4$.

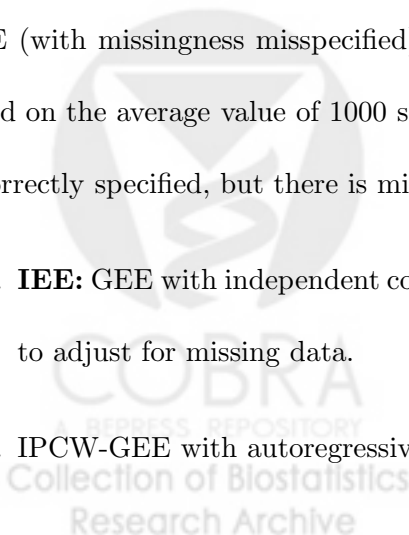
Monotone dropout is generated by a logit-linear selection model with group-by-response inter-

actions. Dropout patterns are based on *level* and *increment* in the response.⁵ Level-based patterns (in which the cumulative effect of responses affects the dropout rate) may be less dependent on unobserved values than dropout that depends on increments, or transitions, between adjacent response values. *Main-effects* dropout models depend on values of the response. For *group-by-response* dropout models, the control group dropout depends on response values but the treatment group dropout remains MCAR. This could occur if the treatment has a benign effect unrelated to the response. For example, if a treatment for high blood pressure had mood-enhancing effects, treatment group patients may be equally likely to continue, regardless of whether they experienced angina. In the control group, though, improvement or declining health may affect participation.

Large-sample ($N = 500$) relative efficiency of IPCW-GEE compared to the second-order MTM is reported as the ratio of sample variances of regression parameter estimates. Regression, MTM association, and IPCW-GEE missingness models are correctly specified, so that regression estimates are consistent for MAR data. IPCW-GEE models are fitted using independence and autoregressive working correlations. Additional simulations explore the role of cluster size in efficiency.

We evaluate bias under MAR dropout for the MTM (with association misspecified) and IPCW-GEE (with missingness misspecified). Relative bias as a percentage of true value β is computed based on the average value of 1000 sets of parameter estimates $\hat{\beta}_j$. In all cases the marginal mean is correctly specified, but there is misspecification elsewhere for the following fitted models:

1. **IEE:** GEE with independent correlation structure (a generalized linear model) and no weights to adjust for missing data.
2. IPCW-GEE with autoregressive (AR1) working correlation and underfitted missingness:



- (a) **IPW1:** Selection depends on Y_{t-1} . $(\phi_2, \phi_3, \phi_4) \equiv 0$ in Table 1.
- (b) **IPW2:** Selection depends on Y_{t-1} and Y_{t-2} . $(\phi_3, \phi_4) \equiv 0$.
- (c) **IPW3:** Selection depends on a group-by-response interaction for Y_{t-1} only. $(\phi_2, \phi_4) \equiv 0$.

3. **MTM(1):** First-order MTM. The true association is second-order, so the likelihood is not correctly specified.

Note that IPW2 is correctly specified for main-effects dropout models in Table 1. Also, all IPCW-GEE models are correctly specified (or overspecified) for *Current* patterns with response main effects, and the IPW3 selection model is correct for the Current pattern with group-by-response interaction.

Under nonignorable (NI) dropout, IPCW-GEE with censoring weights based on MAR missingness will not provide consistent estimates of regression parameters. MTM regression parameters are also expected to be biased if the dropout model is not correctly specified. For NI dropout scenarios, we examine how percent bias varies for different regression coefficients, missingness models, and amounts of missing data. Bias is estimated for the following fitted models:

1. **IEE:** GEE with independent correlation structure (a generalized linear model).
2. IPCW-GEE
 - (a) **IPW2:** Selection depends on Y_{it-1} and Y_{it-2} . Compare to main-effects NI dropout, dependent on Y_{it} and Y_{it-1} .
 - (b) **IPW4:** Selection depends on Y_{it-1} and Y_{it-2} , plus interactions with treatment group. Compare to NI group-by-response dropout.

3. **MAR:** Second-order MTM (assumes MAR data).
4. **NI:** Second-order MTM with misspecified selection model for nonignorable missingness:
 - (a) **NI1:** Selection model estimates ϕ_0 and ϕ_2 (and ϕ_4 for group-by-response selection), and holds λ (values for ϕ_1, ϕ_3) constant as $(-\frac{\phi_1}{2}, -\frac{\phi_3}{2})$.
 - (b) **NI2:** Similar to NI1, but with λ held constant as $(\frac{\phi_1}{2}, \frac{\phi_3}{2})$.

Additional simulations with 50% missing data and the *Level* selection pattern explore whether bias is relative (a percentage of regression parameter values) or absolute (on a log-odds ratio scale). Fitzmaurice, Molenberghs and Lipsitz³² examined asymptotic bias of a range of regression parameter values, holding association and percent missing constant. They found evidence that bias is absolute, not relative. Regression parameter vector $\beta = (1.2, -2.5, 0.65, -0.31)$ is chosen to double the **group** parameter value (the primary parameter of interest for treatment differences) while maintaining trajectories with similar shape to the original model in Table 1.

3.1. Results: Efficiency

When 20% of data are missing, efficiency of IPCW-GEE regression covariates is generally over 90% compared to the MTM. However, efficiency loss for IPCW-GEE is substantial when 50% of data are missing: Figure 3 summarizes IPCW-GEE efficiency under main effects and group-by-response dropout. Percent efficiencies for different covariates in the same model are connected by solid lines (for cluster size 5 in complete data) or dashed lines (cluster size 10).

Unlike prior research,¹⁶ we do not find GEE efficiency greater for cluster-level covariates. For models with a main-effects dropout model, **intercept** has slightly greater efficiency than other

covariates, while models with a group-by-response dropout model show greater efficiency for covariates `group` and `group:time`.

IPCW-GEE estimators exhibit higher efficiency under a more complicated missingness model. Figure 3 shows that efficiency for `group` and `group:time` under group-by-response selection models is greater than for the same covariates in the simpler main-effects selection model. A similar counterintuitive trend is seen for missingness patterns. The *Current* selection model has fewer parameters to fit, and lower efficiency in all regression covariates compared to *Level* and *Increment*.

Comparing dashed and solid lines in Figure 3, efficiency is greater for all parameters when cluster size is doubled. This is probably due to weaker marginal correlations for second-order dependence in larger clusters.¹⁶ The simulations show a small efficiency advantage for independence working correlation over autoregression (not shown). This was unexpected. Previous research shows GEE or IPCW-GEE to be somewhat more efficient with working correlation closer to the true association.^{12,16,33,21} The efficiency differences are slight, usually less than 5 percentage points compared to the MTM.

3.2. Results: Bias

As discussed above, bias in marginal regression parameters is expected for likelihood-based methods when data are MAR and the association model is incorrect, or when data are NI and association or missingness is incorrectly specified. IPCW-GEE may yield biased estimates when the missingness model is incorrect.



3.2.1. MAR main-effects dropout

In six MAR main-effects selection model scenarios (Level, Current, and Increment for 20% or 50% missing data), fitted IPCW-GEE and MTM regression parameters show up to -22% bias. However, bias is negligible (less than $\pm 5\%$) in the `group` and `group:time` parameters that would be of primary interest in a clinical trial. The naive IEE model shows up to -125% bias in the intercept, but only up to -5% bias in `group` and `group:time`.

3.2.2. MAR group-by-response dropout

Figure 4 shows bias of fitted regression parameters for data generated by the group-by-response selection model with 50% missing. Trends in bias for 20% missing data are similar, with about half the magnitude in percent bias. The four panels display results for the four marginal mean covariates. Each panel displays three missingness patterns (Level, Increment, and Current response). Error bars reflect one standard deviation in each direction for fitted parameters in the simulations. A vertical dashed line represents the true value of each covariate, and vertical dotted lines show percent bias in 50% increments. Bias is negative (toward the null value) in all but a few cases.

For all parameters, MTM(1) bias is relatively small. Bias in `intercept` and `time` is somewhat smaller for Model IPW1 than for IEE. This suggests that underspecified selection weights are better than no selection weights. IPW2 (not shown), with second-order response main effects in the missingness model, does not mitigate bias much more than IPW1, but a first-order selection model with group-by-response interaction (IPW3) shows greatly reduced bias in `group` and `group:time` compared to IPW1.

IEE bias in `group` and `group:time` may be smaller than for IPW1 because half the data (the

control group) have MCAR data, as assumed by IEE. (See Table 1.) IPW1 and IPW2 assume that both groups follow the same missingness trend, in which prior symptom expression makes dropout more likely. Fitted values based on average parameter estimates are unbiased for all models at baseline (`time` = -4), but over time fitted IPW1 and IPW2 response probabilities are biased downward for the control group and upward for the treatment group. This dilutes the group difference at `time` = 0 (an end-of-study odds ratio of 3.5), estimated as 2.3 for the IPW1 Level model (3.2 by IPW3, 3.3 by MTM(1)).

3.2.3. NI main-effects dropout

Patterns of bias when missingness is nonignorable (NI) are similar to those of the MAR models: IEE models generally show the most bias, and MTMs assuming MAR can show less bias than the IPCW-GEE. However, magnitude of bias is generally larger for NI models. Simulations with main-effects selection (not shown) reveal bias in the intercept as great as -407%, but even IEE models show little bias in parameters `group` and `group:time`.

3.2.4. NI group-by-response dropout

Figure 5 shows percent bias with 50% NI missingness and a group-by-response interaction in the selection model. Model IPW4 is an IPCW-GEE model with missingness an MAR version (modeling on Y_{it-1} and Y_{it-2}) of the true selection (based on Y_{it} and Y_{it-1}). Bias in IPW4 is somewhat smaller than for marginalized transition model assuming MAR (MTM-MAR) for all missingness patterns. Both generally show less bias than IEE estimates, with the *Increment* model as an exception.

MTM model NI1, with parameters ϕ_1 and ϕ_3 from Table 1 fixed in the wrong direction, does lead to greater bias than for MTM-MAR. Fitting a MTM with NI missingness and (ϕ_1, ϕ_3) at half their true values (in the correct direction) leads to some mitigation of bias. For most models and parameters, percent bias in Model NI2 is about half that of Model MTM-MAR.

3.2.5. Absolute vs. relative bias

When the group parameter is doubled, visual inspection (not shown) suggests that percent bias for the group parameter decreases, while absolute bias between models is relatively constant. Like Fitzmaurice et al.,³² we conclude that bias is probably of less concern the farther a parameter is from a null value. Absolute bias appears to be the underlying bias mechanism. However, we continue to report relative bias, because of greater interpretability for parameters in the same regression model, and for different regression models.

4. ANALYSIS OF PANSS SCHIZOPHRENIA DATA

We return to the PANSS data introduced in Section 1, to model the longitudinal trajectory of schizophrenia symptoms recorded at 5 visits over 8 weeks. We apply the marginalized transition model (MTM) and compare results to models fitted using independence estimating equations (IEE) and inverse probability of censoring weighted generalized estimating equations (IPCW-GEE). We also examine sensitivity of MTM parameter estimates to assumptions about missing at random (MAR) and nonignorable (NI) dropout processes. Treatment assignments for the 420 subjects are: 85 placebo, 85 haloperidol, and 250 risperidone (6 mg, 10 mg, or 16 mg/day). The 420 subjects contribute 1609 response measurements: 491 (23%) overall are missing due to monotone dropout.

4.1. PANSS Fitted Models

The MTM and IPCW-GEE both require specification of three components: a marginal regression model, an association model, and a dropout model. Table 2 summarizes the models used for analysis of the schizophrenia data. The marginal regression model reflects the primary research question: how do patients respond to risperidone, placebo, and haloperidol in the treatment of schizophrenia? Level of improvement and both linear and quadratic time trends are allowed to vary by treatment group. Covariate **week** takes values (-7, -6, -4, -2, 0) so that variables **PLAC** and **RISP** are interpretable as comparisons to haloperidol after 8 weeks of treatment.

Two IPCW-GEE censoring weight models are computed, with and without interactions between treatment group and the first-order lagged response. Four MTMs with NI dropout are fitted, to address sensitivity of MTM parameter estimates to hypothesized missing data mechanisms. Holding other dropout parameters constant, when $\lambda_0 = 1.5$ (Table 2, Model NI1) the odds of remaining in the study when clinical improvement has occurred are $\exp(1.5) \approx 4.5$ times the odds if clinical improvement has not occurred. For $\lambda_0 = -1.5$ (Model NI2) the effect is reversed. The odds ratio of retention by improvement is greater in Model NI3 for the risperidone group than for other groups, and lesser in Model NI4.

4.2. PANSS Marginal Mean Results

Figure 6 shows fitted trajectories for some of the models described in Table 2, and for a generalized linear model estimated by GEE with independence working correlation (IEE). Time trends for risperidone and haloperidol have similar shapes, while the placebo group shows a more gradual increase in improvement probabilities over the course of the trial.

Using IEE, the probability of improvement is overestimated for timepoints after baseline, since patients who do not show improvement are more likely to leave the study. By contrast, likelihood-based MTMs implicitly impute the time trend for these dropout cases,²⁶ and IPCW-GEE similarly weights observations based on selection probabilities. However, treatment group differences at $\text{time} = 0$ are similar for IEE and the other MAR models. The odds ratio for risperidone compared to haloperidol is estimated as 1.69 by IPW2 and MTM(2), and 1.60 by the IEE model. Risperidone is also estimated to be superior to placebo at 8 weeks, with odds ratios of 2.03, 1.85, and 1.53 estimated for MTM(2), IPW2, and IEE respectively. However, treatment differences are not statistically significant (for .05 Type I error) in any of the fitted MAR models. In contrast, a *last observation carried forward* (LOCF) analysis comparing week 8 values to baseline finds risperidone to be superior to both placebo (odds ratio=3.9, $Z=4.05$, $p < .001$) and haloperidol (odds ratio=1.9, $Z=2.33$, $p < .01$). LOCF fails to adjust for changing rates of improvement over time, and for differential dropout rates among the three treatment groups.

Predicted improvement probabilities are higher overall for Model NI2 than for models assuming MAR, due to implicit imputation of responses for healthier subjects. Differential dropout models for the treatment group widens treatment group comparisons in the NI4 fitted marginal mean, and narrows them in Model NI3. The sensitivity analysis does affect inference: NI4 is the only model to find a statistically significant difference between risperidone and haloperidol after 8 weeks of treatment (odds ratio=2.03, $Z=2.10$, $p = .04$), or overall ($\chi^2_3 = 8.8$, $p = .03$). Risperidone and placebo differ significantly at 8 weeks in NI1 (odds ratio 2.39, $Z=2.31$, $p = .02$) and NI4 (odds ratio=2.65, $Z=2.56$, $p = .01$). We conclude that risperidone is not clearly superior to placebo or haloperidol. The scientific conclusions are not robust to assumptions about dropout.

To further justify that a single selection model is not definitive for the PANSS data, Figure 7 displays the MTM profile likelihood for a range of values for $\lambda = (\lambda_0, \lambda_1)$, as defined in Table 2. Contour lines show 50%, 90%, and 95% confidence regions based on $(2 \cdot \log \text{likelihood})$. The maximized profile likelihood is marked by an X, and MAR ($\lambda_0 = \lambda_1 = 0$) is shown to be within the 90% and 95% confidence regions.

The profile likelihood is relatively flat, with a large range of plausible parameter values within a 95% confidence region surrounding the weakly identified maximum of $\lambda = (-0.5, -0.5)$. For example, the 95% confidence region includes both $\lambda = (-1.7, 0)$ and $\lambda = (0.4, 0)$. These sample models differ with respect to both fitted retention probabilities and (nonsignificant) treatment effects: haloperidol appears inferior to placebo at 8 weeks when $\lambda_0 = -1.7$ (OR = 0.90), and superior when $\lambda_0 = 0.4$ (OR = 1.26).

The profile likelihood surface is useful in guiding the choice of selection model scenarios to consider for sensitivity analysis. Certain configurations within a joint 95% confidence region, such as $\lambda = (-1, -1)$ or Model NI2, may be ruled out since such parameters values are clearly not supported by the observed data (see Figure 1) and the adopted model. We are fairly certain that people who show improvement are *not* more likely to drop out at that time. An additional caveat regarding interpretation of Figure 7 is that fitted selection model parameters ϕ do not remain constant as λ values are set at different values. For example, ϕ_4 (the coefficient for Y_{t-1}) decreases nearly linearly from 2.28 to 0.30 as λ_0 increases from -2.0 to 2.0 (while $\lambda_1 = 0$). Although fitted ϕ parameters do compensate somewhat for extreme fixed values of λ , fitted probabilities for selection (and fitted regression model parameters) still vary greatly for different choices of λ . We again state a preference for sensitivity analysis over estimation of weakly identified parameters: “It’s not what

you know that hurts you; it's the things you think you know but don't".¹¹

4.3. PANSS Serial Dependence and Dropout Results

The MTM(2) fitted association model (not shown) suggests that the odds of improvement given improvement at the previous timepoint are over 10 times the odds of improvement having not shown improvement at the previous assessment. Additionally, the odds of improvement are 3.75 times as great if the pre-previous response shows improvement since baseline. The MAR dropout model used for Model IPW1 depends on the previous two response values and treatment group. Patients who show clinical improvement in two consecutive measurements have odds of staying in the study for the next measurement approximately 2.5 times the odds for patients in the same treatment group who had not shown clinical improvement at either time. This *Level* effect suggests that people whose symptoms improve compared to baseline are more likely to stay in the study. The *Increment* effect is measured as odds of retention at time t for a subject who does not show improvement at time $t - 2$ but does at time $t - 1$: the odds of retention are about 10.1 times higher than the odds of retention for someone who shows improvement at time $t - 2$ but not at time $t - 1$. People may be more likely to stay as the treatment seems to start working. Level and increment effects are quite different in the NI models, since some parameters are held fixed.

In summary, the MTM accounting for dropout is a rich model that permits maximum likelihood estimates for clinically important, interpretable parameters. For the PANSS data, in which dropout appears to vary by both treatment group and response values, treatment effects are attenuated compared to semiparametric models that do not accommodate MAR dropout. Sensitivity analysis with four hypothesized models for nonignorable dropout changes inference for regression model

parameters. Fitted values for MTM models with nonignorable dropout vary somewhat more than for MTM and IPCW-GEE models assuming data are MAR.

5. DISCUSSION

For models where the focus of interest is the marginal mean, $\mu_{it}^M = \text{logit}^{-1}(X_{it}\beta)$, correct specification of the regression model will result in valid inference for β if data are MCAR.¹² When data are MAR, the likelihood-based marginalized transition model (MTM) requires specification of the conditional mean as well, via a transition model. Inverse probability of censoring weighted GEE (IPCW-GEE),¹³ a semiparametric method, does not require correct specification of an association model under MAR. However, a correctly specified dropout model must be used as weights in the estimating equation for β . Under MAR, the modeling burden for MTM and IPCW-GEE is similar. The MTM models current response given prior response values, while IPCW-GEE models current dropout status given prior responses. The simulation studies in Section 3 demonstrate that comparable misspecification leads to comparable bias. When missingness is nonignorable, the MTM likelihood for estimation of β must also include a properly specified dropout model. Both MTM and IPCW-GEE depend on sensitivity analysis to examine validity of inference in the presence of nonignorable dropout.¹⁴

For models assuming MAR dropout, comparing misspecification bias of a likelihood-based MTM (misspecified serial association) to IPCW-GEE (misspecified dropout) may seem like comparing apples and oranges. MTM association and IPCW-GEE dropout are both logit-linear models dependent on prior responses, but their misspecification cannot strictly be compared. However, the analyst can explore and compare both methods applied to the observed data. Depending on cluster

length, sampling schemes, and diversity of reasons for dropout, the MTM or IPCW-GEE may be selected as a primary analysis method. The MTM has the additional advantages of efficiency and availability of likelihood ratio tests for comparing nested models.

Doubly robust estimation³⁴ could yield consistent estimation of GEE regression parameters even if the MAR or NI dropout model is incorrect. For doubly robust estimation, the GEE quasi-score is augmented by a function of the responses. If the augmentation is selected correctly and modeled according to the complete data distribution, regression parameters will be consistent regardless of dropout model. Conversely, if the dropout model is correct, the augmentation does not need to be correctly specified in order for regression parameters to be consistent. Despite the added modeling burden of the augmented quasi-score, doubly robust estimation has promise in minimizing the impact of NI dropout. However, the augmented quasi-score is difficult to define and solve, so doubly robust estimation for modeling longitudinal responses is sometimes viewed as impractical.³⁵

The IPCW-GEE models described show great efficiency loss compared to the MTM, but IPCW-GEE as described here is not the most efficient semiparametric estimation method for β .³⁶ However, an adaptive estimator for the efficient score function for β requires several additional finite-parameter regression models (Equations (11), (12), and (13) of Robins and Rotnitzky³⁶). These three models add considerable overhead to the modeling process. IPCW-GEE requires only the marginal regression model (to estimate parameters of interest), working correlation (including independence), and a dropout model that adds complexity linearly by cluster length and model order. By contrast, the models required for adaptive estimates of efficient IPCW-GEE scores add $\binom{n_i}{2}$ models in addition to estimating (β, ϕ) . The semiparametric efficiency bound is not attained if these additional models are misspecified. Although consistent estimation of regression parameters

β is robust to misspecification of the additional models for the adaptive estimator, the modeling burden is equivalent to that of *both* weighting methods such as IPCW-GEE and imputation such as the sequential imputation of Paik.²⁰

The MTM for NI dropout is a likelihood-based method incorporating a marginal regression model with association and dropout models that depend on response values. As presented here, the method does not accommodate intermittent missing data or continuous time Markov dependence. These are areas for future work.

REFERENCES

1. Chouinard, G., Jones, B., Remington, G., Bloom, D., Addington, D., MacEwan, G., Labelle, A., Beaulclair, L. and Arnott, W. 'A Canadian multicenter placebo-controlled study of fixed doses of risperidone and haloperidol in the treatment of chronic schizophrenia patients', *Journal of Clinical Psychopharmacology*, **13**, 25–40 (1993).
2. Kay, S. R., Fiszbein, A. and Opler, L. A. 'The positive and negative syndrome scale (PANSS) for schizophrenia', *Schizophrenia Bulletin*, **13**, 261–276 (1987).
3. Diggle, P. J., Heagerty, P. J., Liang, K.-Y. and Zeger, S. L. *Analysis of Longitudinal Data*, second edn, Clarendon Press, Oxford, 2002.
4. Marder, S. R. and Meibach, R. C. 'Risperidone in the treatment of schizophrenia', *The American Journal of Psychiatry*, **151**, 825–835 (1994).
5. Diggle, P. and Kenward, M. G. 'Informative drop-out in longitudinal data analysis (with discussion)', *Applied Statistics*, **43**, 49–93 (1994).
6. Little, R. J. A. 'Pattern-mixture models for multivariate incomplete data', *Journal of the American Statistical Association*, **88**, 125–134 (1993).

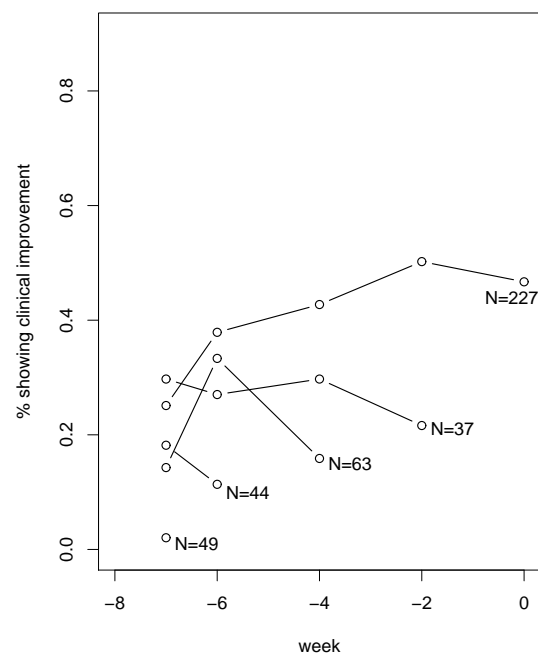


Figure 1: PANSS schizophrenia data: percent showing clinically significant improvement over time, for completers (N=227) and for patients who drop out after 1, 2, 4, or 6 weeks of treatment.

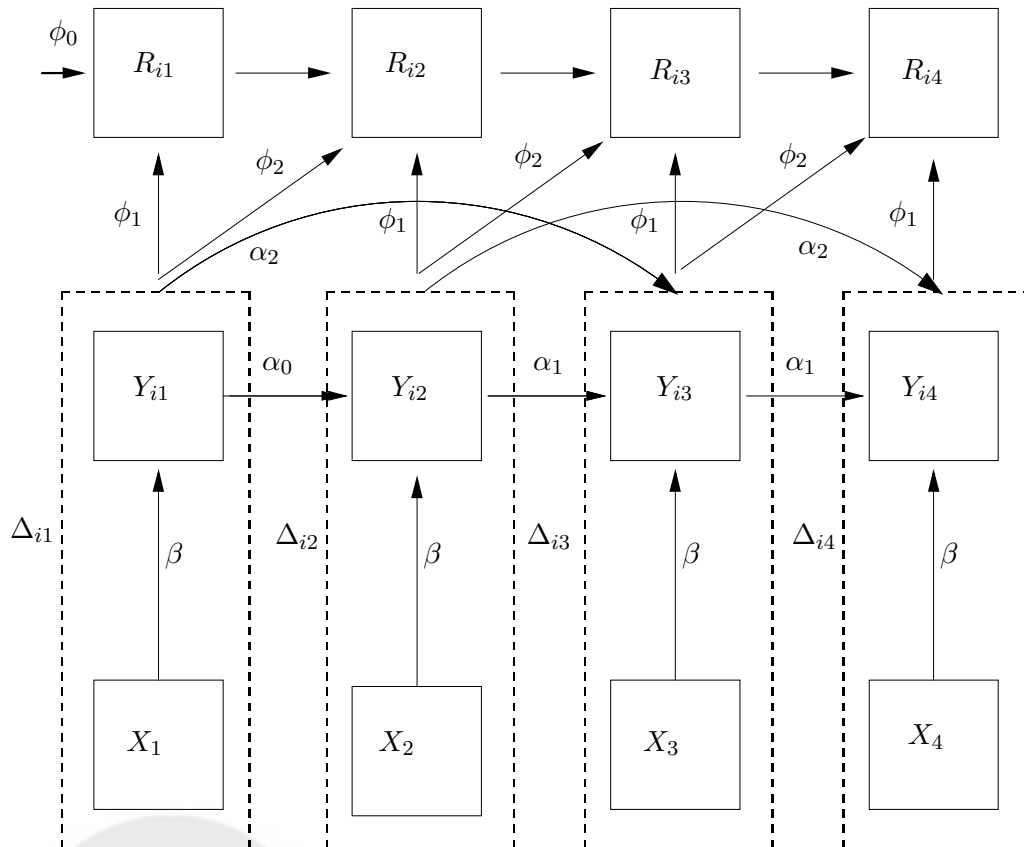


Figure 2: Schematic of a second-order marginalized transition model with nonignorable monotone dropout. Illustrates the marginal model (X, β) , conditional model (Δ, α) , and dropout model (ϕ) .

Table 1: Data generation for efficiency and misspecification bias simulations.

Marginal Mean

$$\text{logit} [P(Y_{it} = 1|X_{it})] = \beta_0 + \beta_1 \cdot \text{group}_i + \beta_2 \cdot \text{time}_{it} + \beta_3 \cdot \text{group}_i \cdot \text{time}_{it}$$

$$\beta = (0.40, -1.25, 0.65, -0.31); \quad \text{time} = (-4, -3, -2, -1, 0)$$

Serial Dependence

$$\text{logit} [P(Y_{it} = 1|X_{it}, Y_{it-1}, Y_{it-2})] = \Delta_{it} + \alpha_1 \cdot y_{it-1} + \alpha_2 \cdot y_{it-2}$$

$$\alpha = (\log(8), \log(2)); \quad \alpha_2 = 0 \text{ for } t = 1 \text{ (time} = -4)$$

Dropout

$$\text{logit} \left[P(R_{it} = 1 | R_{it-1} = 1, Y_{it}, H_{it}^{(Y)}, X_{it}) \right] =$$

$$\text{(MAR)} \quad \phi_0 + \phi_1 \cdot y_{it-1} + \phi_2 \cdot y_{it-2} + \phi_3 \cdot \text{group}_i \cdot y_{it-1} + \phi_4 \cdot \text{group}_i \cdot y_{it-2}$$

$$\text{(NI)} \quad \phi_0 + \phi_1 \cdot y_{it} + \phi_2 \cdot y_{it-1} + \phi_3 \cdot \text{group}_i \cdot y_{it} + \phi_4 \cdot \text{group}_i \cdot y_{it-1}$$

main effects

group-by-response

$(\phi_1, \phi_2, \phi_3, \phi_4)$

$(\phi_1, \phi_2, \phi_3, \phi_4)$

Level $(-0.75, -0.75, 0, 0)$

$(-0.75, -0.75, 0.75, 0.75)$

Current $(-1.5, 0, 0, 0)$

$(-1.5, 0, 1.5, 0)$

Increment $(-0.75, 0.75, 0, 0)$

$(-0.75, 0.75, 0.75, -0.75)$

ϕ_0 set for 20%, 50% missing data overall

Other Details

sample size = 500 (250 per group)

1000 simulated datasets per scenario

cluster size = 5 for complete data, except as noted

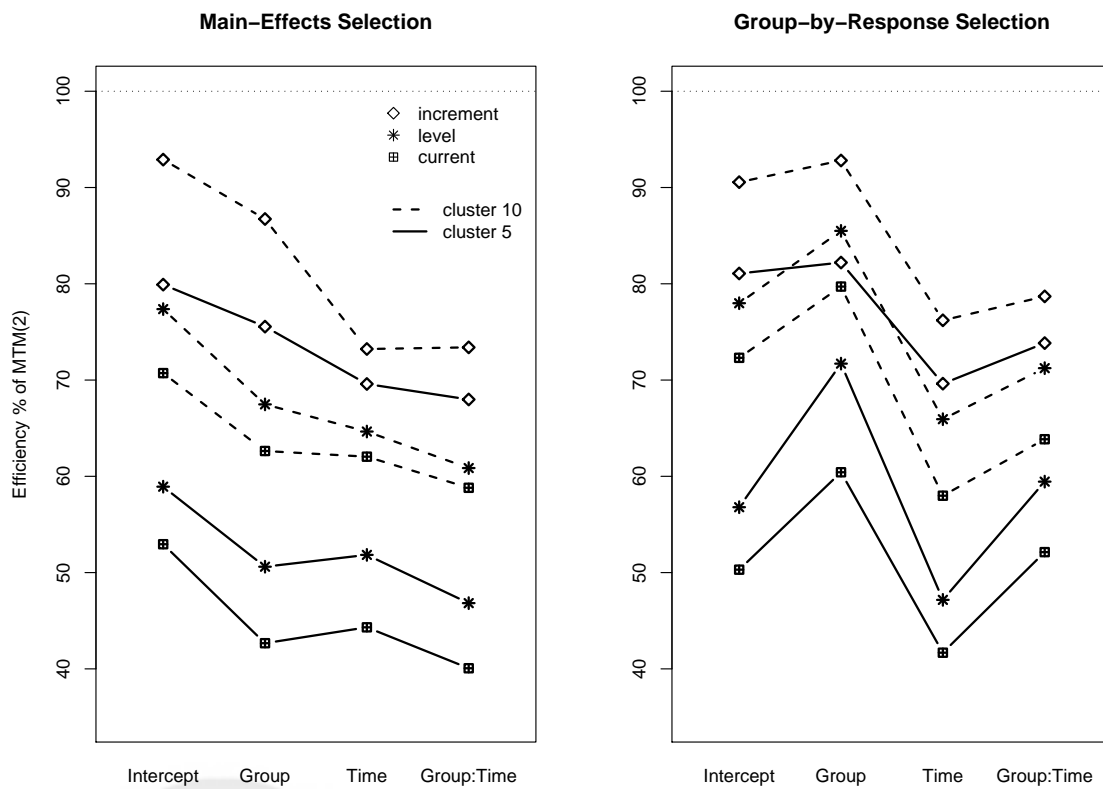


Figure 3: Efficiency of IPCW-GEE compared to MTM(2) for correctly specified regression and dropout models (MAR, 50% missing).

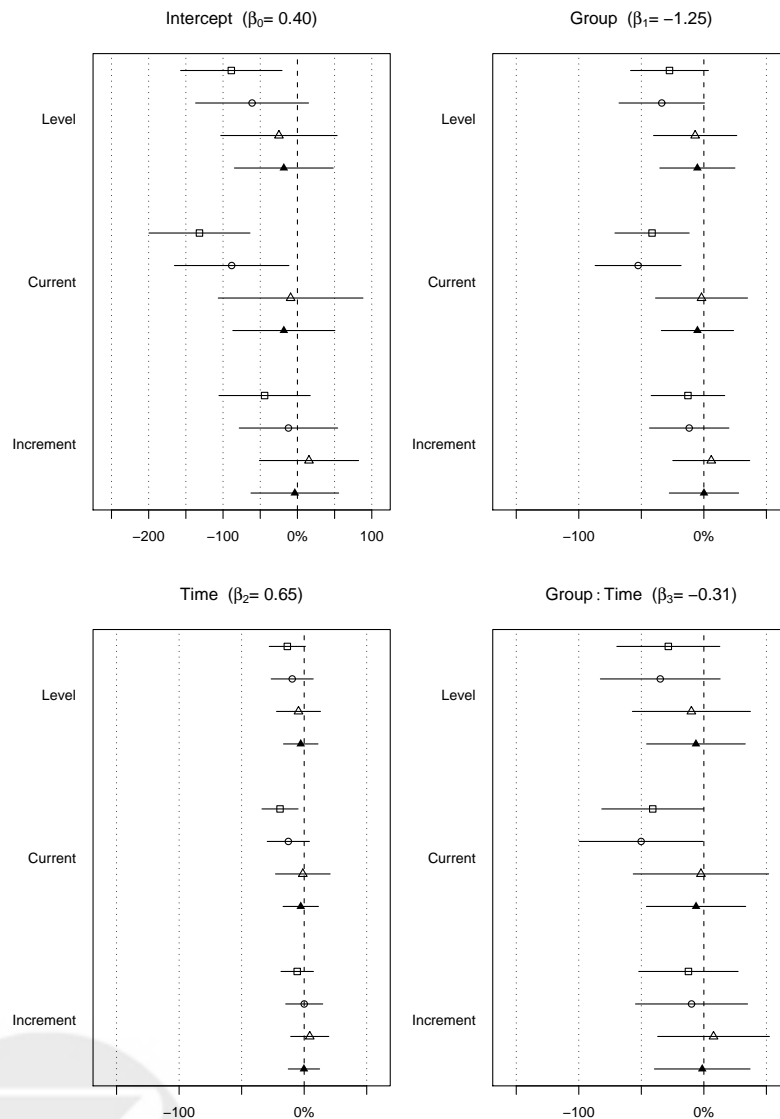


Figure 4: Average regression parameter estimates \pm one standard deviation based on 1000 simulations, for the marginal mean model in Table 1. 50% of data are missing at random (MAR), with selection dependent on a group-by-response interaction.

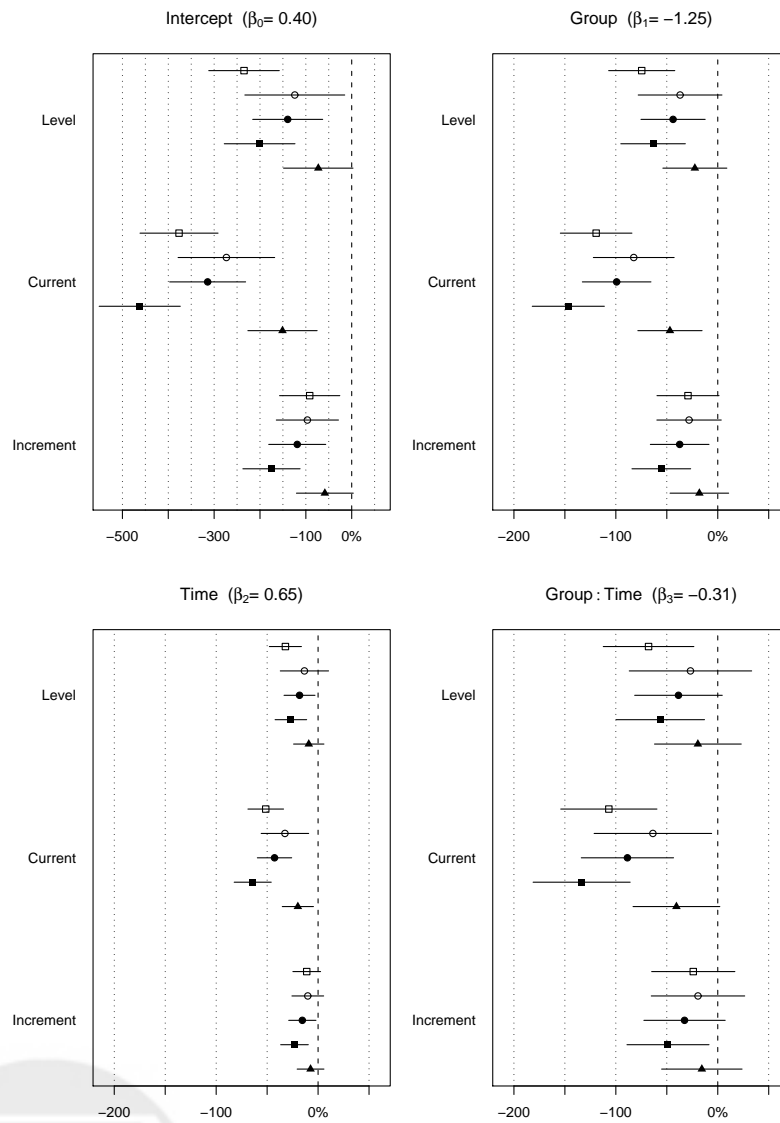


Figure 5: Average regression parameter estimates \pm one standard deviation based on 1000 simulations, for the marginal mean model in Table 1. Missingness (50%) is dependent on current response (NI) through a group-by-response interaction.

Table 2: Summary of models considered for schizophrenia data.

Marginal Mean (all models)

$$\begin{aligned} \text{logit}[P(Y_{it} = 1|X_{it})] = & \beta_0 + \beta_1 \cdot \text{PLAC}_i + \beta_2 \cdot \text{RISP}_i + \beta_3 \cdot \text{week}_{it} + \beta_4 \cdot \text{week}_{it}^2 \\ & + \beta_5 \cdot \text{PLAC}_i \cdot \text{week}_{it} + \beta_6 \cdot \text{PLAC}_i \cdot \text{week}_{it}^2 + \beta_7 \cdot \text{RISP}_i \cdot \text{week}_{it} + \beta_8 \cdot \text{RISP}_i \cdot \text{week}_{it}^2 \end{aligned}$$

Serial Dependence

Marginalized Transition Models:

$$\begin{aligned} \text{logit}[P(Y_{it} = 1|X_{it}, Y_{it-1}, Y_{it-2})] = \\ t = 2 \quad \quad \quad \Delta_{i2}^{(2)} + \alpha_0 \cdot Y_{it-1} \\ t > 2 \quad \quad \quad \Delta_{it} + \alpha_1 \cdot Y_{it-1} + \alpha_2 \cdot Y_{it-2} \end{aligned}$$

IPCW-GEE: Autoregressive working correlation

Dropout

$$\begin{aligned} \text{logit}\left[P(R_{it} = 1|R_{it-1} = 1, Y_{it}, H_{it}^{(Y)}, X_{it})\right] = \\ t = 2 \quad \quad \quad \phi_0 \\ t > 2 \\ \text{(IPW1)} \quad \quad \quad \phi_1 + \phi_2 \cdot \text{PLAC}_i + \phi_3 \cdot \text{RISP}_i + \phi_4 \cdot y_{it-1} + \phi_5 \cdot y_{it-2} \\ \text{(IPW2)} \quad \quad \quad \phi_1 + \phi_2 \cdot \text{PLAC}_i + \phi_3 \cdot \text{RISP}_i + \phi_4 \cdot y_{it-1} + \phi_5 \cdot y_{it-2} \\ \quad \quad \quad + \phi_6 \cdot y_{it-1} \cdot \text{PLAC}_i + \phi_7 \cdot y_{it-1} \cdot \text{RISP}_i \\ \text{(NI1: } \lambda_0 = 1.5, \lambda_1 = 0) \quad \phi_1 + \phi_2 \cdot \text{PLAC}_i + \phi_3 \cdot \text{RISP}_i + \phi_4 \cdot y_{it-1} + \lambda_0 \cdot y_{it} + \lambda_1 \cdot y_{it} \cdot \text{RISP}_i \\ \text{(NI2: } \lambda_0 = -1.5, \lambda_1 = 0) \\ \text{(NI3: } \lambda_0 = 1.5, \lambda_1 = 1) \\ \text{(NI4: } \lambda_0 = 0.75, \lambda_1 = -1) \end{aligned}$$

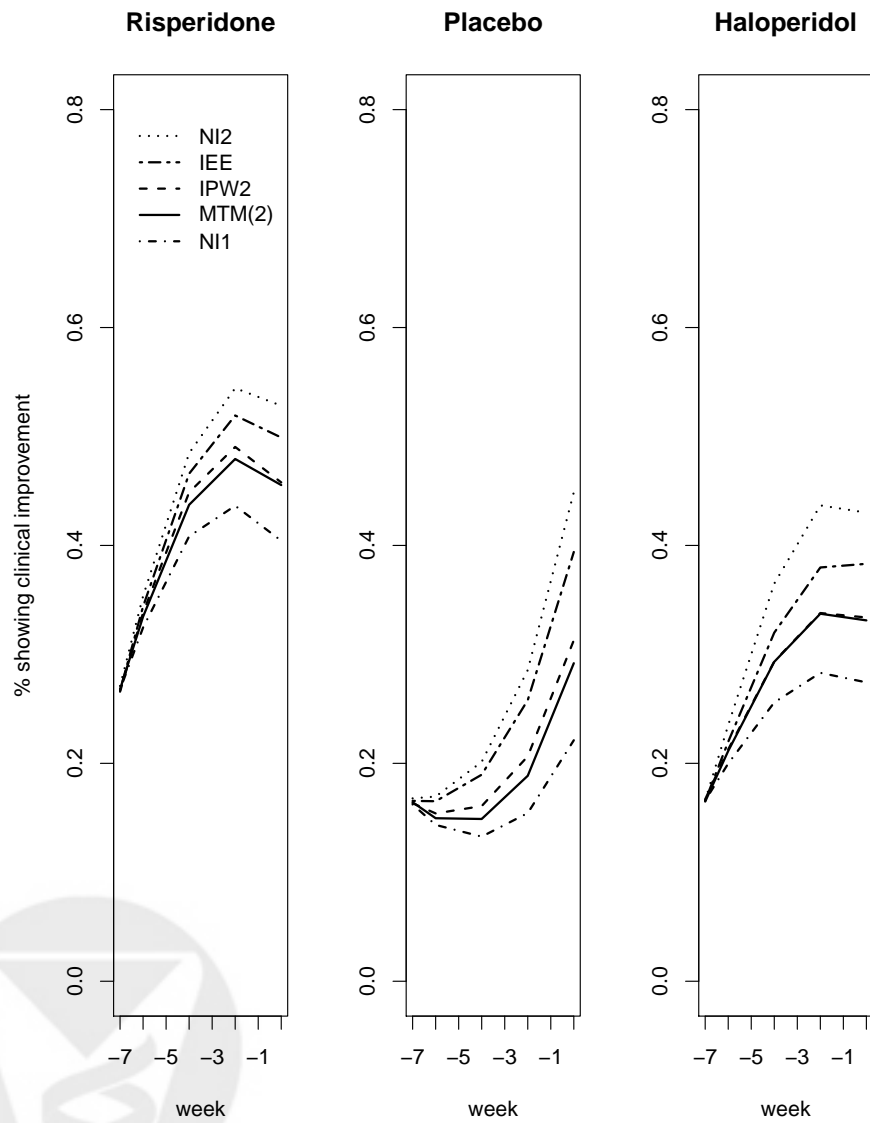


Figure 6: Fitted trajectories for schizophrenia data (N=420).

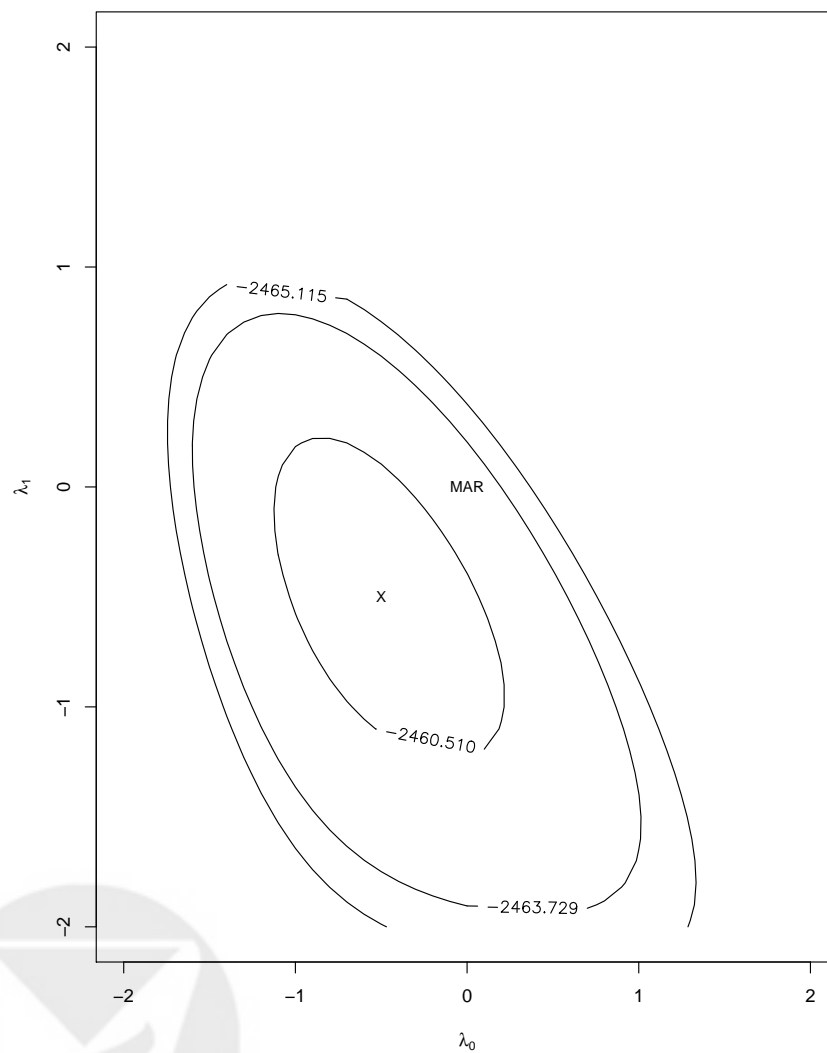
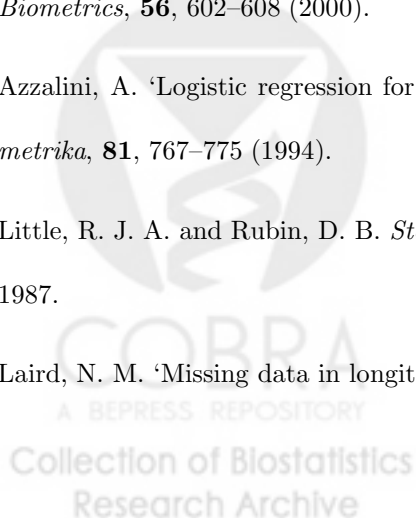


Figure 7: NI 2-log likelihood surface for schizophrenia data ($N=420$), with contours for 50, 90, 95% confidence regions.

7. Wu, M. C. and Carroll, R. J. 'Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process (Corr: V45 p1347; V47 p357)', *Biometrics*, **44**, 175–188 (1988).
8. Zeger, S. L., Liang, K.-Y. and Albert, P. S. 'Models for longitudinal data: A generalized estimating equations approach', *Biometrics*, **44**, 1049–1060 (1988).
9. Fitzmaurice, G. M., Laird, N. M. and Zahner, G. E. P. 'Multivariate logistic models for incomplete binary responses', *Journal of the American Statistical Association*, **91**, 99–108 (1996).
10. Molenberghs, G., Kenward, M. G. and Lesaffre, E. 'The analysis of longitudinal ordinal data with nonrandom drop-out', *Biometrika*, **84**(1), 33–44 (1997).
11. Scharfstein, D. O., Rotnitzky, A. and Robins, J. M. 'Adjusting for nonignorable drop-out using semiparametric nonresponse models (with discussion)', *Journal of the American Statistical Association*, **94**, 1096–1146 (1999).
12. Liang, K.-Y. and Zeger, S. L. 'Longitudinal data analysis using generalized linear models', *Biometrika*, **73**, 13–22 (1986).
13. Robins, J. M., Rotnitzky, A. and Zhao, L. P. 'Analysis of semiparametric regression models for repeated outcomes in the presence of missing data', *Journal of the American Statistical Association*, **90**, 106–121 (1995).
14. Rotnitzky, A., Robins, J. M. and Scharfstein, D. O. 'Semiparametric regression for repeated outcomes with nonignorable nonresponse', *Journal of the American Statistical Association*, **93**(444), 1321–1339 (1998).
15. Preisser, J. S., Galecki, A. T., Lohman, K. K. and Wagenknecht, L. E. 'Analysis of smoking trends with incomplete longitudinal binary responses', *Journal of the American Statistical Association*, **95**, 1021–1031 (2000).

16. Fitzmaurice, G. M., Laird, N. M. and Rotnitzky, A. G. 'Regression models for discrete longitudinal response (with discussion)', *Statistical Science*, **8**, 284–309 (1993).
17. Rubin, D. B. *Multiple imputation for nonresponse in surveys*, John Wiley & Sons, New York, 1987.
18. Cowles, M. K., Carlin, B. P. and Connett, J. E. 'Bayesian tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness', *Journal of the American Statistical Association*, **91**, 86–98 (1996).
19. Collins, L. M., Schafer, J. L. and Kam, C.-M. 'A comparison of inclusive and restrictive strategies in modern missing data procedures', *Psychological Methods*, **6**, 330–351 (2002).
20. Paik, M. C. 'The generalized estimating equation approach when data are not missing completely at random', *Journal of the American Statistical Association*, **92**(440), 1320–1329 (1997).
21. Heagerty, P. J. 'Marginalized transition models and likelihood inference for longitudinal categorical data', *Biometrics*, **58**, 342–351 (2002).
22. Heagerty, P. J. and Zeger, S. L. 'Marginalized multilevel models and likelihood inference (with discussion)', *Statistical Science*, **15**, 1–26 (2000).
23. Albert, P. S. 'A transitional model for longitudinal binary data subject to nonignorable missing data', *Biometrics*, **56**, 602–608 (2000).
24. Azzalini, A. 'Logistic regression for autocorrelated data with application to repeated measures', *Biometrika*, **81**, 767–775 (1994).
25. Little, R. J. A. and Rubin, D. B. *Statistical analysis with missing data*, John Wiley & Sons, New York, 1987.
26. Laird, N. M. 'Missing data in longitudinal studies', *Statistics in Medicine*, **7**, 305–315 (1988).



27. Willett, J. B. and Singer, J. D. 'Investigating onset, cessation, relapse, and recovery: Why you should, and how you can, use discrete-time survival analysis to examine event occurrence', *Journal of Consulting and Clinical Psychology*, **61**, 952–965 (1993).
28. Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. *Numerical Recipes in C: The Art of Scientific Computing*, second edn, Cambridge University Press, Cambridge, UK, 1992.
29. Kurland, B. F. Analysis of binary longitudinal data with dropout and death, PhD thesis, University of Washington Department of Biostatistics, Seattle, WA, 2002.
30. Fitzmaurice, G. M., Heath, A. F. and Clifford, P. 'Logistic regression models for binary panel data with attrition', *Journal of the Royal Statistical Society, Series A*, **159**, 249–263 (1996).
31. Preisser, J. S., Lohman, K. K. and Rathouz, P. J. 'Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random', *Statistics in Medicine*, **21**, 3035–3054 (2002).
32. Fitzmaurice, G. M., Molenberghs, G. and Lipsitz, S. R. 'Regression models for longitudinal binary responses with informative drop-outs', *Journal of the Royal Statistical Society, Series B*, **57**, 691–704 (1995).
33. Mancl, L. A. and Leroux, B. G. 'Efficiency of regression estimates for clustered data', *Biometrics*, **52**, 500–511 (1996).
34. Robins, J. M. and Rotnitzky, A. 'On double robustness', *Statistica Sinica*, **11**(4), 920–936 (2001).
35. Scharfstein, D. O. and Irizarry, R. A. 'Generalized additive selection models for the analysis of studies with potentially nonignorable missing outcome data', *Biometrics*, **59**, 601–613 (2003).
36. Robins, J. M. and Rotnitzky, A. 'Semiparametric efficiency in multivariate regression models with missing data', *Journal of the American Statistical Association*, **90**, 122–129 (1995).